

A Convergence Theory for a Class of
Quasi-Newton Methods for Constrained Optimization¹

by

Rodrigo Fontecilla²

Trond Steihaug³

Richard A. Tapia⁴

Technical Report 83-15, May 1983.
(revised February 1986)

¹This work was supported by DOE contract DE-ASO5-82-ER13016, ARO DAAG-29-83-K-0035 and the General Research Board of the University of Maryland. An earlier draft of this paper was presented at the SIAM National Meeting, Denver, Colorado, June 1983.

²Department of Computer Science, and the Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742.

³Department of Mathematical Sciences, Rice University, Houston, Texas. Present address: Statoil, 4001 Stavanger, Norway.

⁴Department of Mathematical Sciences, Rice University, P.O. Box 1892, Houston, Texas 77251.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE FEB 1986		2. REPORT TYPE		3. DATES COVERED 00-00-1986 to 00-00-1986	
4. TITLE AND SUBTITLE A Convergence Theory for a Class of Quasi-Newton Methods for Constrained Optimization				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Computer Science, Institute for Physical Science and Technology, University of Maryland, College Park, MD, 20742				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT see report					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 28	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

A convergence theory for a class of quasi-Newton methods for constrained optimization¹

*Rodrigo Fontecilla*²

*Trond Steihaug*³

*Richard A. Tapia*⁴

Abstract. In this paper we develop a general convergence theory for a class of quasi-Newton methods for equality constrained optimization. The theory is set in the framework of the diagonalized multiplier method defined by Tapia and is an extension of the theory developed by Glad. We believe that this framework is flexible and amenable to convergence analysis and generalizations. A key ingredient of a method in this class is a multiplier update. Our theory is tested by showing that a straightforward application gives the best known convergence results for several known multiplier updates. Also a characterization of q-superlinear convergence is presented. It is shown that in the special case when the diagonalized multiplier method is equivalent to the successive quadratic programming approach, our general characterization result gives the Boggs, Tolle and Wang characterization.

1. Introduction. This paper considers a class of quasi-Newton methods for solving the equality constrained minimization problem:

$$\text{minimize } f(x) \tag{1.1}$$

$$\text{subject to } g(x) = 0$$

where $f : R^n \rightarrow R$ and $g : R^n \rightarrow R^m$.

The augmented Lagrangian $L : R^n \times R^m \times R_+ \rightarrow R$ is given by

$$L(x, \lambda, c) = f(x) + g(x)^t \lambda + \frac{c}{2} g(x)^t g(x).$$

¹This work was partially supported by DOE contract DE-ASO5-82ERI 13016, ARO DAAG-29-83-K-0035 and the Venezuelan government. This paper was presented at the SIAM National Meeting held at Denver, Colorado, June 1983.

² Department of Computer Science, and the Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742.

³ Statoil, 4001 Stavanger, Norway.

⁴ Mathematical Sciences Department, Rice University, P.O.Box 1892, Houston, Texas 77251.

For c equal to zero, the augmented Lagrangian reduces to the standard Lagrangian

$$l(x, \lambda) = f(x) + g(x)^t \lambda.$$

If $x_* \in R^n$ is such that $\nabla g(x_*)$ is full rank, then a necessary condition for x_* to be a solution of (1.1) is that there exists $\lambda_* \in R^m$ such that (x_*, λ_*) is a solution of the non-linear system

$$\nabla_x L(x_*, \lambda_*, c) = 0, \quad (1.2.a)$$

$$g(x_*) = 0. \quad (1.2.b)$$

Moreover, in this case λ_* will be unique. It may be noted that the constant c does not affect condition (1.2).

In order to approximate the minimizer x_* we consider the Diagonalized Multiplier Method (DMM), as defined by Tapia [39].

Given x_0, λ_0, B_0 .

For $k = 0$ Until convergence Do

$$\lambda_{k+1} = U(x_k, \lambda_k, B_k) \quad (1.3)$$

$$B_k s_k = -\nabla_x L(x_k, \lambda_{k+1}, c) \quad (1.4)$$

$$x_{k+1} = x_k + s_k \quad (1.5)$$

$$B_{k+1} = B(x_k, \lambda_k, B_k). \quad (1.6)$$

The matrices B_k in (1.4) and B_{k+1} in (1.6) are intended to be approximations to the Hessian matrix $\nabla_x^2 L(x_*, \lambda_*, c)$. We call U in (1.3) a multiplier update and B in (1.6) an approximate Hessian update. Implicit in the formulation of the DMM is the option of changing U or B at each iteration.

On occasions in the DMM we will refer to a particular choice of quasi-Newton method for the steps (1.4)-(1.6). For example, diagonalize Newton multiplier method would mean that the choice for B_{k+1} in (1.6) is $\nabla_x^2 L(x_{k+1}, \lambda_{k+1}, c)$; while diagonalized secant multiplier method would emphasize that the quasi-Newton method (1.4)-(1.6) is also a secant method.

If B in (1.6) is a secant update (see Dennis and Schnabel [18] for details on these methods) then the default choice for y_k (the structure of the problem does not suggest a more natural choice) is the choice given by Tapia [39]; namely

$$y_k = \nabla_x L(x_{k+1}, \lambda_{k+1}, c) - \nabla_x L(x_k, \lambda_{k+1}, c). \quad (1.7)$$

Recall that a secant update requires the satisfaction of the secant equation

$$B_{k+1} s_k = y_k. \quad (1.8)$$

In the original formulation of the multiplier method (given independently by Hestenes [28], Powell [33] and Haarhoff and Buys [8]) the multiplier λ_k was updated only after x_k had been found which minimized $L(x, \lambda_k, c)$ in x (satisfied $\nabla_x L(x_k, \lambda_k, c) = 0$). The process of finding such an x_k was left undefined. Suppose that these unconstrained minimizations were performed via a quasi-Newton method. Then in terms of (1.3)-(1.6) the multiplier method amounts to looping through (1.4)-(1.6) an infinite number of times before returning to the multiplier update (1.3). Of course, in practice such an approach would be impossible and only a finite number of loops of (1.4)-(1.6) (quasi-Newton steps) could be taken before returning to the multiplier update (1.3). Tapia [39] formally stated (1.3)-(1.6) and used the adjective diagonalized to describe this modified version of the multiplier method which gave the multiplier the same status as the x -variable. He was motivated by the feeling that in an effective formulation the multiplier λ should be updated as often as the variable x and the two update formulas should be matched or compatible in some sense.

The following multiplier updates are well-known and appear throughout the literature:

$$U(x, \lambda, B) = \lambda + c \ g(x) \quad (1.9)$$

$$U(x, \lambda, B) = -(\nabla g^t(x) \nabla g(x))^{-1} \nabla g^t(x) \nabla f(x) \quad (1.10)$$

$$U(x, \lambda, B) = \lambda + (\nabla g^t(x) B^{-1} \nabla g(x))^{-1} g(x) \quad (1.11)$$

$$U(x, \lambda, B) = \lambda + (\nabla g^t(x) B^{-1} \nabla g(x))^{-1} (g(x) - \nabla g^t(x) B^{-1} \nabla_x L(x, \lambda, c)) \quad (1.12)$$

We call (1.9) the *Hestenes-Powell update* since it was the update proposed independently by both Hestenes [28] and Powell [33] when they introduced the multiplier method. Haarhoff and Buys [8] used (1.10) with their version of the multiplier method. However, it had appeared in the literature numerous times before it was used by them. We call it the *projection update* since it can be obtained as the least squares solution for λ of the linear system $\nabla_x L(x, \lambda) = 0$. The update (1.11) is due to Buys, more will be said about it later on. Following Tapia [39,40] we refer to (1.2) as the *extended problem*, since it involves both x and λ as unknowns. Fletcher [20] calls Newton's method on the extended problem the *Solver method*. It is well-known that the diagonalized Newton multiplier method using the multiplier update (1.12) is equivalent to Newton's method on the extended problem. It is this equivalence which motivated us to call the multiplier update (1.12) the Newton multiplier update. For a background on the multiplier method and related issues see Bertsekas [3].

In the remainder of this introductory section we will accomplish three objectives. Firstly, we will motivate the choice of the DMM as the framework for our unified theory. Secondly, we will present a fairly complete historical account of the development of the convergence theory for quasi-Newton methods for constrained optimization as it relates to the theory developed in this paper. This historical account will give a prospective to our contribution in terms of existing results, it will lend support to our reasons for favoring the DMM formulation and finally it is needed in its own right since the field has advanced significantly in the last 10 years and there is considerable confusion as to the particular contributions of the various authors. Thirdly, we will briefly describe what the reader will encounter in the remaining sections of this paper.

We now motivate our choice for the DMM framework. The DMM using the Newton multiplier update (1.12) is equivalent to numerous quasi-Newton formulations for problem (1.1) (see Tapia [40]). These equivalent formulations include a structured quasi-Newton method on the extended problem, the popular successive quadratic programming (SQP) quasi-Newton method, and a formulation which Tapia calls the structured multiplier substitution method. The role of the multiplier update is most prominent in the DMM. It is less prominent in the extended problem formulation, even less prominent in the SQP formulation and essentially masked in the structured multiplier substitution method.

Since the DMM formulation clearly delineates the role of the multiplier update, when coupled with a good convergence theory it should allow one to determine exactly what properties a multiplier update must satisfy for a particular application or result. This is the path that was taken by Fontecilla [21] using the theory developed in this paper.

The Newton multiplier update (1.12) is the only multiplier update which causes the DMM to satisfy linearized constraints (see Theorem 10.2 of Tapia [40] and also Fontecilla [21]). Since the other formulations described above, including the SQP formulation, satisfy linearized constraints it follows that the DMM offers a broader framework than do the other frameworks. Namely, it allows one to consider algorithms which do not necessarily satisfy linearized constraints and when linearized constraints are satisfied it gives an equivalent formulation. We believe that contemporary approaches designed with global behavior in mind will not necessarily satisfy linearized constraints. This is certainly true of the trust region algorithm for constrained optimization recently suggested by Celis, Dennis and Tapia [11]. Their algorithm can be described in the DMM framework. For an interesting class of algorithms which generalize the DMM see Fontecilla [22].

In terms of popularity there is no doubt that the SQP formulation has won over the DMM formulation. However, in terms of amenability to convergence analysis and generalization we feel that the DMM may offer distinct advantages.

We would now like to present a fairly complete historical account of the development of the convergence theory for quasi-Newton methods for constrained optimization as it relates to the present work. Convergence theory for algorithms that use an approximation to the projected Hessian is not of the same flavor as that presented here and will not be discussed. The reader interested in projected Hessian quasi-Newton methods is referred to the recent papers by Coleman and Conn [14], Nocedal and Overton [31], Fontecilla [22] and Byrd [10].

We are concerned with convergence results for quasi-Newton methods for constrained optimization which work with an approximation to the full Hessian with respect to x . Our discussion will center on the following papers: Buys [7], Garcia-Palomares and Mangasarian [23], Han [26], Tapia [39], Powell [34], Byrd [9], Glad [25] and Boggs, Tolle and Wang [5]. While our choice of papers is not exhaustive it is much more than representative and should give a good prospective to the theory presented in this paper.

To begin with the extended problem is a part of the folklore of constrained optimization theory. The standard convergence theory for Newton's method can be used to establish local quadratic convergence in (x, λ) of Newton's method applied to the extended problem. Furthermore, the standard Broyden, Dennis Moré [6] convergence theory for secant methods can be used to establish local q-superlinear convergence in

(x, λ) of a standard secant method applied to the extended problem as long as this particular secant method does not require positive definiteness of the matrix that is being approximated, e.g. Broyden or PSB (see Dennis and Schnabel [18] for more details on secant methods). These facts were all well-known at the time the popular secant updates for nonlinear equations and unconstrained optimization were being developed. However, it was generally felt that the extended problem approach was unsatisfactory. Firstly because there was no underlying unconstrained minimization problem to use for guidance and secondly the popular DFP and BFGS secant methods were precluded due to the fact that the Jacobian of the full system (Hessian of the augmented Lagrangian with respect to both x and λ) while being symmetric, is necessarily not positive definite. We mention in passing that it is ironic that many authors taking directions away from the extended problem (e.g. SQP or DMM approach) either openly or tacitly returned to it for their convergence analysis. At any rate the stage was set for considerable research activity to shift to the multiplier methods as soon as they were introduced. After all, they did contain a fundamental unconstrained minimization problem.

In an enlightening thesis Buys [7] showed that the multiplier method using the Hestenes-Powell multiplier update (1.9) is the gradient method with step length parameter c on the dual problem. He then proposed the multiplier update (1.11) for use with the multiplier method since the resulting algorithm would be Newton's method on the dual problem. Convergence results followed from standard theory for these two forms of the multiplier method.

Tapia [39] formally defined the diagonalized multiplier method (DMM) and demonstrated local q-superlinear convergence in (x, λ) of several diagonalized secant multiplier methods using the Newton multiplier update (1.12) including the DFP and BFGS secant methods. Various algorithms which could be classified as diagonalized multiplier methods had previously appeared in the literature. For example, several authors including Bard and Greenstadt [2] and Tapia [37,38] considered algorithms which were essentially the diagonalized Newton multiplier method using the Newton multiplier update (1.12). That their algorithm was equivalent to Newton's method on the extended problem was known to Bard and Greenstadt [2] and to Tapia [38] but not in [37]. Miele, Cragg, Iyer and Levy [29] had previously proposed the diagonalized gradient multiplier method using the Hestenes-Powell multiplier update (1.9). They gave no convergence analysis but included a considerable amount of numerical experimentation.

Byrd [9] considered a generalization of the diagonalized Newton multiplier method where j_k Newton steps were taken on the unconstrained minimization problem before λ_k was updated to λ_{k+1} . He proved, among other things, the interesting result that a multiplier update, e.g. (1.11), gives local q-quadratic convergence for the multiplier method if and only if this multiplier update gives local q-quadratic convergence in (x, λ) in his modified form of the DMM for any choice of j_k satisfying $j_k \geq 2$. Namely, two Newton steps on the unconstrained minimization subproblem are sufficient to obtain the optimal quadratic convergence rate in (x, λ) .

Byrd's result coupled with the known fact that the diagonalized Newton multiplier method using the Newton multiplier update gave local q-quadratic convergence in (x, λ) essentially removed the multiplier method (i.e. any implementation that required a large number of quasi-Newton steps in the unconstrained minimization phase) from consideration as an effective algorithm and gave further impetus to the DMM.

About the same time that the DMM was emerging as a viable formulation, the SQP approach was surfacing as an attractive and viable formulation for quasi-Newton methods for constrained optimization. Garcia-Palomares and Mangasarian [23] following Wilson [41], who had presented us with the SQP Newton method (exact Hessian was used), proposed an SQP quasi-Newton method where the approximation to the Hessian used in the quadratic term was taken as the upper left-hand $n \times n$ submatrix of a quasi-Newton approximation to the $(n+m) \times (n+m)$ Jacobian of the extended problem, i.e., the $(n+m) \times (n+m)$ Hessian with respect to (x, λ) of the (augmented) Lagrangian. They established various r-convergence results in (x, λ) .

Han [26,27] improved and polished the Garcia-Palomares and Mangasarian formulation by using a secant method to directly approximate the Hessian with respect to x of the (augmented) Lagrangian, i.e., the $n \times n$ submatrix referred above, and presented us with the SQP secant methods as we know them today. He established local q-superlinear convergence in (x, λ) for numerous secant updates including the DFP and the BFGS.

Glad [25] independently also defined the formal DMM. He established local convergence results for the diagonalized BFGS secant multiplier method using the Hestenes-Powell (1.9), the projection (1.10) and the Newton (1.12) multiplier updates. Specifically, he obtained local q-linear convergence in (x, λ) for the Hestenes-Powell update, local q-linear convergence in x for the projection update, and local q-superlinear convergence in (x, λ) for the Newton update. While the results for the Newton update had previously been obtained by Tapia[39], Glad obtained them independently. To our knowledge Glad [25] was the first to give any convergence results for a secant method for constrained optimization which was not equivalent to the SQP secant method. His work contributed significantly to our understanding of the DMM.

All convergence results mentioned above for the DFP or the BFGS secant update either carried with them the assumption that the Hessian with respect to x of the Lagrangian was positive definite or the author worked with the augmented Lagrangian and assumed that c was sufficiently large so that the Hessian with respect to x of the augmented Lagrangian was positive definite near the solution. Furthermore, Han, Tapia and Glad all used the Broyden-Dennis-Moré convergence theory for secant methods and all performed their convergence analysis using a form of the extended problem. It is not surprising then that their results are essentially the same and in particular they all obtained q-superlinear convergence in (x, λ) for the diagonalized secant multiplier methods using the Newton multiplier update or the equivalent SQP secant methods.

Tapia [39,40] considered the convergence rate given by these algorithms for the x variable alone. Clearly, in general a q-rate in (x, λ) implies no more than the corresponding r-rate in x (or in λ). He observed that if in the approximation formula used for the Hessian (1.6) λ was replaced by a multiplier estimate which did not depend on λ , e.g. the projection update (1.10), then the q-superlinear convergence rate also applied to the variable x alone. Glad [25] also observed that if the multiplier update used in the DMM did not depend on λ or c , then the convergence result could be stated in x alone. Powell [34] obtained an r-superlinear convergence rate for his modified form of the SQP BFGS secant method. He expressed concern over the fact that he had not been able to obtain a q-superlinear rate in x ; but he seemed not to realize that up to this time no one had obtained a q-superlinear rate in x for the SQP BFGS secant method or any similar algorithm. In the same paper Powell derived a condition which implied 2-step q-superlinear convergence in x for an SQP quasi-Newton method. This result fueled the already burning interest in extending the well-known Dennis-Moré [17] characterization of q-

superlinear convergence of quasi-Newton methods for unconstrained optimization to a characterization of those quasi-Newton updates which when used in the SQP quasi-Newton method gave q-superlinear convergence in x .

Boggs, Tolle and Wang [5] answered both of above open questions. Namely, working with the DMM and the Newton multiplier update (1.12) they derived a characterization of those approximate Hessian updates which led to q-superlinear convergence in x . They then demonstrated the usefulness of their characterization by using it to prove that the DFP and BFGS secant updates (assuming positive definiteness of the Hessian with respect to x at the solution) gave q-superlinear convergence in x without any modifications as had previously been suggested by Tapia [39].

The work of Han [27], Tapia [39,40], Glad [25] and Boggs, Tolle and Wang [5] has greatly influenced the present work. We now describe what the reader will encounter in the material of the paper and relate this material to the existing work that has just been described. In Section 2, we list the basic assumptions and standard lemmas that will be used in the remainder of the paper. We give a formal definition of the notion of bounded deterioration of approximate Hessian updates used in the DMM. We also describe various properties which a multiplier update may possess. These properties will allow us to determine various convergence results based on the theory developed in Section 3.

In Section 3 we follow the Broyden, Dennis and Moré [6] convergence theory and develop a unified convergence theory for the DMM. This theory requires that B in (1.6) be of bounded deterioration in the sense defined in Section 2. The convergence theory allows us to determine if the use of a particular multiplier update will give local q-linear convergence in (x, λ) or the stronger result of local q-linear convergence in x alone.

In Section 4 we apply the tools developed in Sections 2 and 3 to the standard multiplier updates (1.9)-(1.12). We do this as a test and demonstration of the unified theory. We do not wish to imply that these standard multiplier updates should be used; instead we feel that the understanding gained from these demonstrations may be beneficial in the design and analysis of new algorithms. In each case we see that the unified theory gives results which are as good or better than those that presently exist in the literature. The convergence result for the Buys (1.11) multiplier update is new (Proposition 4.4). It is satisfying that our theory (Proposition 4.2) for the Newton multiplier update (equivalently SQP) matches the Boggs, Tolle and Wang [5] convergence result and is superior to the convergence results given by Han [26], Tapia [39] and Glad [25].

In Section 5 we derive two characterizations of those update pairs (U, B) where U is a multiplier update and B is an approximate Hessian update which lead to q-superlinear convergence in x . These characterizations are Theorems 5.1 and 5.3. We then show that if U is the Newton update (1.12), then Corollary 5.4 gives the Boggs, Tolle and Wang characterization. However, we obtain the result under slightly less restrictive assumptions than Boggs, Tolle and Wang used. Recently Nocedal and Overton [31] have also obtained the Boggs, Tolle and Wang characterization under these less restrictive assumptions. We emphasize that our two characterization results are for the general DMM and in the special case that the Newton multiplier update is used we obtain the Boggs, Tolle and Wang characterization.

2. Preliminaries. Recall that x_* is a local solution of problem (1.1) with associated multiplier λ_* . To simplify our notation we let

$$\nabla g(x_k) = \nabla g_k, \quad \text{and} \quad \nabla g(x_*) = \nabla g_*. \quad (2.1.a)$$

$$A^c = \nabla_x^2 L(x_*, \lambda_*, c) \quad (2.1.b)$$

$$A_* = \nabla_x^2 l(x_*, \lambda_*). \quad (2.1.c)$$

Throughout this paper we will be making the following assumptions:

A1. The functions f and g have second derivatives which are Lipschitz continuous in an open neighborhood D of x_* .

A2. ∇g_* has full rank.

A3. $z^t A_* z > 0$ for all $z \neq 0$ satisfying $\nabla g_*^t z = 0$.

A4. A_* is nonsingular.

Assumption A3 is the well-known second order sufficiency condition from constrained optimization. Moreover, it can be shown (Buys [7]) that A2 and A3 are equivalent to asking that $\nabla^2 l(x_*, \lambda_*)$ be nonsingular. Assumption A1 and the nonsingularity of $\nabla^2 l(x_*, \lambda_*)$ are the standard assumptions made when considering the convergence theory for quasi-Newton methods on the extended problem; and as such are minimal assumptions. Many of the results that follow could be proved without assuming A4. However, the generality lost by assuming A4 is not of major concern here.

The following lemmas will play a fundamental role in the analysis presented in the remaining sections.

LEMMA 2.1. *There exists $\bar{c} \geq 0$ such that A^c is positive definite for all $c \geq \bar{c}$. Moreover, for $c > \bar{c}$, letting A denote $A^{\bar{c}}$, A^c denote A^c and ∇g denote ∇g_* , we have*

$$(A^c)^{-1} = A^{-1} - A^{-1} \nabla g \left[(c - \bar{c})^{-1} I + \nabla g^t A^{-1} \nabla g \right]^{-1} \nabla g^t A^{-1} \quad (2.2.a)$$

and

$$\nabla g^t (A^c)^{-1} \nabla g = (c - \bar{c})^{-1} \nabla g^t A^{-1} \nabla g \left[(c - \bar{c})^{-1} I + \nabla g^t A^{-1} \nabla g \right]^{-1}; \quad (2.2.b)$$

so that

$$(A^c)^{-1} \rightarrow \left[I - A^{-1} \nabla g (\nabla g^t A^{-1} \nabla g)^{-1} \nabla g^t \right] A^{-1}, \quad (2.2.c)$$

$$c \nabla g^t (A^c)^{-1} \nabla g \rightarrow I, \text{ and} \quad (2.2.d)$$

$$(A^c)^{-1} \nabla g \rightarrow 0 \quad \text{as } c \rightarrow \infty. \quad (2.2.e)$$

Proof. The first statement of this lemma is standard (see Lemma 1.25 of Bertsekas [3]). The statements (2.2.d) and (2.2.e) were known to Glad [25] when he briefly suggested a method of proof. Below we have expanded Glad's suggested proof to the point where it can be followed with only a fair amount of effort.

Start by writing

$$A^c = A + (c - \bar{c})\nabla g \nabla g^t. \quad (2.3)$$

By setting A equal to our A , $U=(c-\bar{c})\nabla g$ and $V=\nabla g$ in equation (13) on page 50 of Ortega and Rheinboldt [32] (Sherman-Morrison-Woodbury formula) we obtain (2.2.a). Multiply (2.2.a) on the left by ∇g^t and on the right by ∇g and take the first expression of the form $\nabla g^t A^{-1} \nabla g$ and write it as $\nabla g^t A^{-1} \nabla g B^{-1} B$ where $B=(c-\bar{c})^{-1}I + \nabla g^t A^{-1} \nabla g$. Now by factoring out $\nabla g^t A^{-1} \nabla g B^{-1}$ from both terms we arrive at (2.2.b). The expressions (2.2.c)-(2.2.e) are direct consequences of (2.2.a) and (2.2.b). \square

In this paper we use $|\cdot|$ to denote both the l_2 vector-norm and the matrix norm that it induces and we use $\|\cdot\|$ to denote an arbitrary but fixed matrix norm. However, since all norms in a finite dimensional space are equivalent we know that for the norm $\|\cdot\|$ there exist $\mu, \eta > 0$ such that for all $A \in R^{n \times n}$ we have

$$\mu \|A\| \leq |A| \leq \eta \|A\|. \quad (2.4)$$

The following lemma can be found in Dennis and Schnabel [18].

LEMMA 2.2. Assume $F: R^n \rightarrow R^n$ is differentiable in the open convex set D , and suppose that for some w^* in D and all $w \in D$

$$|F'(w) - F'(w^*)| \leq K |w - w^*|. \quad (2.5)$$

for a positive constant K . Then for each u and v in D ,

$$|F(v) - F(u) - F'(w^*)(v - u)| \leq K \max\{|v - w^*|, |u - w^*|\} |v - u|. \quad (2.6)$$

Moreover, if $F'(w^*)$ is invertible, then there is an $\epsilon > 0$, $\alpha > 0$ and $\beta > 0$ such that $\max\{|v - w^*|, |u - w^*|\} \leq \epsilon$ implies that u and v belong to D and

$$\alpha |v - u| \leq |F(u) - F(v)| \leq \beta |v - u|. \quad (2.7)$$

The following lemma was first stated formally by Han [26]. It was used implicitly by Tapia [39,40], Glad [25] and Boggs, Tolle and Wang [5].

LEMMA 2.3. For each fixed value of c there exist positive constants K_1 and K_2 and an $\epsilon(c) > 0$ such that for all $\lambda \in R^m$ and for any u, v satisfying $\sigma(u, v) \leq \epsilon(c)$ we have

$$|\nabla_x L(v, \lambda, c) - \nabla_x L(u, \lambda, c) - A^c(v - u)| \leq [K_1 \sigma(u, v) + K_2 |\lambda - \lambda_*|] |v - u| \quad (2.8)$$

where $\sigma(u, v) = \max\{|v - x_*|, |u - x_*|\}$.

Proof. Let D in A1 play the role of D in Lemma 2.2 and for a fixed c let $\nabla_x L(\cdot, \lambda_*, c)$ play the role of $F(\cdot)$. We know from assumption A1 that there exists $K_1 > 0$ such that

$$|\nabla_x L(v, \lambda_*, c) - \nabla_x L(u, \lambda_*, c)| \leq K_1 |v - u| \quad (2.9)$$

for all $u, v \in D$. A straightforward calculation gives

$$\nabla_x L(v, \lambda, c) - \nabla_x L(u, \lambda, c) - A^c(v - u)$$

$$\begin{aligned}
 &= \nabla_x L(v, \lambda_*, c) - \nabla_x L(u, \lambda_*, c) - A^s(v - u) \\
 &+ [\nabla g(v) - \nabla g(u)](\lambda - \lambda_*). \tag{2.10}
 \end{aligned}$$

Now using the triangle inequality on (2.10) and both (2.6) and (2.7) of Lemma 2.2 we obtain (2.8) with K_2 given by β in (2.7) and $\epsilon(c)$ given by Lemma 2.2. \square

LEMMA 2.4. *There exist positive constants K_3 and K_4 such that for all $v \in R^m$*

$$K_3|v| \geq |\nabla g_* v| \geq K_4|v|. \tag{2.11}$$

Proof. Since ∇g_* is full rank $\nabla g_*^t \nabla g_*$ is nonsingular and

$$v = (\nabla g_*^t \nabla g_*)^{-1} \nabla g_*^t (\nabla g_* v).$$

The result now follows by choosing

$$K_3 = |\nabla g_*| \quad \text{and} \quad K_4 = |(\nabla g_*^t \nabla g_*)^{-1} \nabla g_*^t|^{-1}.$$

\square

The following notion plays a fundamental role in the convergence theory developed in the next section. It is an extension to constrained optimization of the notion of bounded deterioration originated by Dennis [16] and used extensively by Broyden, Dennis and Moré [6].

Consider the pair (U, B) where U is a multiplier update and B is an approximate Hessian update (see (1.3) and (1.6)). Also consider $c \geq 0$ such that A^s is nonsingular. Suppose that U and B are defined in a neighborhood $N = N_1 \times N_2 \times N_3$ of (x_*, λ_*, A^s) where N_3 contains only nonsingular matrices.

DEFINITION 2.5. *The update B is said to be of bounded deterioration (at A^s with respect to U) if there exist non-negative constants α_1 and α_2 such that for each $(x, \lambda, B) \in N$ and for*

$$\lambda_+ = U(x, \lambda, B) \tag{2.12.a}$$

$$x_+ = x - B^{-1} \nabla_x L(x, \lambda_+, c) \tag{2.12.b}$$

$$B_+ = B(x, \lambda, B) \tag{2.12.c}$$

we have

$$\|B_+ - A^s\| \leq \left[1 + \alpha_1 \sigma(z, z_+)\right] \|B - A^s\| + \alpha_2 \sigma(z, z_+) \tag{2.13}$$

with

$$z = (x, \lambda), \quad z_+ = (x_+, \lambda_+) \tag{2.14.a}$$

and

$$\sigma(z, z_+) = \max\{|x - x_*|, |x_+ - x_*|, |\lambda - \lambda_*|, |\lambda_+ - \lambda_*|\}. \tag{2.14.b}$$

Moreover, we say that the multiplier update U is x -dominated (at A^s) if there exists a non-negative constant $\phi(c) < 1$ such that for each $(x, \lambda, B) \in N$ we have

$$|(A^\sharp)^{-1} \nabla g_*(\lambda_+ - \lambda_*)| \leq \phi(c) |x - x_*|. \quad (2.15)$$

Furthermore, we say that the multiplier update U is weakly x -dominated (at A^\sharp with respect to B) if

$$|(A^\sharp)^{-1} \nabla g_*| < 1 \quad (2.16.a)$$

and there exists a non-negative constant $\phi(c) < 1$ such that for each $(x, \lambda, B) \in N$ we have

$$|U(x_+, \lambda_+, B_+) - \lambda_*| \leq \phi(c) \max(|\lambda_+ - \lambda_*|, |x - x_*|). \quad (2.16.b)$$

Finally, we say that the multiplier update U is consistent if it is continuous in N and for all $B \in N_3$ we have

$$\lambda_* = U(x_*, \lambda_*, B). \quad (2.17)$$

Observe that (2.2.e) of Lemma 2.1 says that (2.16.a) will be satisfied for c sufficiently large.

3. Local convergence. This section is devoted to the study of the convergence of the sequences generated by the DMM. Recall assumptions A1 and A2 of Section 2 and Definition 2.5. The proofs of the following two theorems will follow Broyden, Dennis and Moré [6] as closely as possible.

THEOREM 3.1. Consider the update pair (U, B) . Suppose that U is x -dominated at A^\sharp with constant $\phi = \phi(c)$ and B is of bounded deterioration at A^\sharp with respect to U . Then for each $r \in (\phi, 1)$, there exist positive ϵ_x , ϵ_λ and δ such that for

$$|x_0 - x_*| \leq \epsilon_x, \quad |\lambda_0 - \lambda_*| \leq \epsilon_\lambda, \quad \text{and } \|B_0 - A^\sharp\| \leq \delta$$

the sequence $\{(x_k, \lambda_k)\}$ generated by the DMM (1.9)-(1.6) is well defined and converges to (x_*, λ_*) . Furthermore, for all $k \geq 0$ we have

$$|x_{k+1} - x_*| \leq r |x_k - x_*|, \quad (3.1)$$

and $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded.

Proof. Choose positive $\epsilon_x, \epsilon_\lambda$ and δ so that $|x - x_*| \leq \epsilon_x$, $|\lambda - \lambda_*| \leq \epsilon_\lambda$ and $\|B - A^\sharp\| \leq 2\delta$ imply that (x, λ, B) is contained in the neighborhoods qualifying U to be x -dominated and B to be of bounded deterioration. Further, restrict ϵ_x so that $\epsilon_x \leq \epsilon(c)$ where $\epsilon(c)$ is given by Lemma 2.3. Let K_1 and K_2 also be given by Lemma 2.3, let K_4 be given by Lemma 2.4 and let α_1 and α_2 be as in (2.13). For the norm $\|\cdot\|$, let η be given by (2.4). Choose $\gamma_1 \geq |(A^\sharp)^{-1}|$ and $\gamma_2 \geq |A^\sharp|$.

Further restrict $\epsilon_x, \epsilon_\lambda$ and δ so that

$$(2\alpha_1\delta + \alpha_2) \frac{\epsilon_x}{1+r} \leq \delta, \quad (3.2)$$

$$\gamma_1(1+r) \left[(K_1 + K_2 \gamma_2 K_4^{-1} \phi) \epsilon_z + (1+\phi) 2\eta\delta \right] \leq r - \phi, \quad (3.3)$$

and

$$\epsilon_\lambda \leq \min(\epsilon_z, \gamma_2 K_4^{-1} \phi \epsilon_z). \quad (3.4)$$

Now suppose that $|x_o - x_*| \leq \epsilon_z$, $|\lambda_o - \lambda_*| \leq \epsilon_\lambda$ and $\|B_o - A^\sharp\| \leq \delta$. Then $|B_o - A^\sharp| \leq \eta\delta < 2\eta\delta$ and since (3.3) implies that

$$2\gamma_1(1+r)\eta\delta \leq r, \quad (3.5)$$

the Banach Perturbation Lemma [32] gives

$$|B_o^{-1}| \leq (1+r)\gamma_1. \quad (3.6)$$

A straightforward argument gives

$$\begin{aligned} |x_1 - x_*| &\leq |B_o^{-1}| \|\nabla_x L(x_o, \lambda_1, c) - \nabla_x L(x_*, \lambda_1, c) - A^\sharp(x_o - x_*)\| \\ &\quad + |B_o^{-1}| |B_o - A^\sharp| |x_o - x_*| + |B_o^{-1} A^\sharp| |(A^\sharp)^{-1} \nabla g_*(\lambda_1 - \lambda_*)|. \end{aligned} \quad (3.7)$$

The triangle inequality gives

$$|B_o^{-1} A^\sharp| - 1 \leq \|I - B_o^{-1} A^\sharp\| \leq |B_o^{-1}| |B_o - A^\sharp|; \quad (3.8)$$

so it follows that

$$|B_o^{-1} A^\sharp| \leq 1 + (1+r)\gamma_1 2\eta\delta. \quad (3.9)$$

Also,

$$|A^\sharp|^{-1} K_4 |\lambda_1 - \lambda_*| \leq |(A^\sharp)^{-1} \nabla g_*(\lambda_1 - \lambda_*)|; \quad (3.10)$$

so it follows that

$$|\lambda_1 - \lambda_*| \leq \gamma_2 K_4^{-1} \phi |x_o - x_*|. \quad (3.11)$$

Now using Lemma 2.3 with $u = x_o$ and $v = x_*$, (3.9) and (3.11) we obtain from (3.7)

$$|x_1 - x_*| \leq \gamma_1(1+r) \left[(K_1 + K_2 \gamma_2 K_4^{-1} \phi) \epsilon_z + (1+\phi) 2\eta\delta \right] |x_o - x_*| + \phi |x_o - x_*|. \quad (3.12)$$

Using the bound (3.3) we see that (3.12) leads to $|x_1 - x_*| \leq r |x_o - x_*|$. It follows from (3.4) and (3.11) that $|\lambda_1 - \lambda_*| \leq \epsilon_\lambda$.

We complete the proof with an induction argument. Assume that $\|B_k - A^\sharp\| \leq 2\delta$, $|x_{k+1} - x_*| \leq r |x_k - x_*|$ and $|\lambda_{k+1} - \lambda_*| \leq \epsilon_\lambda$ for $k = 1, \dots, m-1$. Observing that (3.4) implies in (2.14.b) that $\sigma(z, z_*) \leq \epsilon_z$, we obtain from (2.13) that

$$\|B_{k+1} - A^\sharp\| - \|B_k - A^\sharp\| \leq 2\alpha_1 \delta \epsilon_z r^k + \alpha_2 \epsilon_z r^k. \quad (3.13)$$

By summing both sides of (3.13) from $k=0$ to $k=m-1$ we obtain

$$\|B_m - A^\sharp\| \leq \|B_0 - A^\sharp\| + (2\alpha_1\delta + \alpha_2) \frac{\epsilon_s}{(1+r)}; \quad (3.14)$$

which by (3.2) implies that $\|B_m - A^\sharp\| \leq 2\delta$. To complete the induction we need to show that $|x_{m+1} - x_*| \leq r|x_m - x_*|$ and $|\lambda_{m+1} - \lambda_*| < \epsilon_\lambda$. These inequalities are established in exactly the way we established them for $m=0$. The boundedness of $\{B_k\}$ and $\{B_k^{-1}\}$ follows directly from the inequalities established above. \square

THEOREM 3.2. *Consider the update pair (U, B) . Suppose that U is consistent and weakly x -dominated at A^\sharp with respect to B with constant $\phi(c)$ and B is of bounded deterioration at A^\sharp with respect to U . Let $\phi = \max(\phi(c), |(A^\sharp)^{-1} \nabla g_*|)$. Then for each $r \in (\phi, 1)$ there exist positive ϵ and δ such that for*

$$|x_0 - x_*| \leq \epsilon, \quad |\lambda_0 - \lambda_*| \leq \epsilon, \quad \text{and} \quad \|B_0 - A^\sharp\| \leq \delta$$

the sequence $\{(x_k, \lambda_k)\}$ generated by the DMM (1.9)-(1.6) is well-defined and converges to (x_*, λ_*) . Furthermore for all $k \geq 0$ we have

$$\max(|x_{k+1} - x_*|, |\lambda_{k+2} - \lambda_*|) \leq r \max(|x_k - x_*|, |\lambda_{k+1} - \lambda_*|) \quad (3.15)$$

and $\{B_k\}$ and $\{B_k^{-1}\}$ are bounded.

Proof. Choose positive ϵ' and δ' so that $|x - x_*| \leq \epsilon'$, $|\lambda - \lambda_*| \leq \epsilon'$ and $\|B - A^\sharp\| \leq 2\delta'$ imply that (x, λ, B) is contained in the neighborhoods qualifying U to be weakly x -dominated and B to be of bounded deterioration. Further restrict ϵ' so that $\epsilon' \leq \epsilon(c)$ where $\epsilon(c)$ is given by Lemma 2.3. Let K_1 and K_2 also be given by Lemma 2.3 and let α_1 and α_2 be as in (2.13). For the norm $\|\cdot\|$, let η be given by (2.4). Choose $\gamma \geq |(A^\sharp)^{-1}|$.

Relying on the consistency of U choose $\epsilon \leq \epsilon'$ and $\delta \leq \delta'$ so that

$$|U(x, \lambda, B) - \lambda_*| \leq \epsilon \quad (3.16)$$

whenever $|x - x_*| \leq \epsilon$, $|\lambda - \lambda_*| \leq \epsilon$ and $\|B - A^\sharp\| \leq \delta$. Further restrict ϵ and δ so that

$$(2\alpha_1\delta + \alpha_2) \frac{\epsilon}{(1+r)} \leq \delta, \quad (3.17)$$

and

$$\gamma(1+r) \left[(K_1 + K_2)\epsilon + 2\eta\delta(1+\phi) \right] \leq r - \phi. \quad (3.18)$$

Now suppose that $|x - x_*| \leq \epsilon$, $|\lambda - \lambda_*| \leq \epsilon$ and $\|B - A^\sharp\| \leq \delta$. Then $|B_0 - A^\sharp| \leq \eta\delta < 2\eta\delta$ and since (3.17) implies that

$$2\gamma(1+r)\eta\delta \leq r, \quad (3.19)$$

the Banach Perturbation Lemma [32] gives

$$|B_0^{-1}| \leq (1+r)\gamma. \quad (3.20)$$

A straightforward argument using (3.7)-(3.9), Lemma 2.3 with $u = x_0$ and $v = x_*$ and

(3.18) gives

$$\begin{aligned}
 |x_1 - x_*| &\leq \gamma(1+r) [K_1\epsilon + 2\eta\delta] |x_0 - x_*| \\
 &\quad + \gamma(1+r) K_2\epsilon |\lambda_1 - \lambda_*| + [1 + (1+r)\gamma 2\eta\delta] \phi |\lambda_1 - \lambda_*| \\
 &\leq \left\{ \gamma(1+r) [(K_1+K_2)\epsilon + 2\eta\delta(1+\phi)] + \phi \right\} \max(|x_0 - x_*|, |\lambda_1 - \lambda_*|) \\
 &\leq r \max(|x_0 - x_*|, |\lambda_1 - \lambda_*|). \tag{3.21}
 \end{aligned}$$

From the definition of weakly x-dominated we have

$$|\lambda_2 - \lambda_*| \leq \phi \max(|x_0 - x_*|, |\lambda_1 - \lambda_*|). \tag{3.22}$$

Finally, (3.21) and (3.22) show that (3.15) and

$$|x_{k+1} - x_*| \leq \epsilon' \quad \text{and} \quad |\lambda_{k+2} - \lambda_*| \leq \epsilon' \tag{3.23}$$

hold for $k=0$. That (3.15) and (3.23) hold for arbitrary $k \geq 0$ can be established by induction in a similar manner. \square

Notice that an x-dominated multiplier update leads to the q-linear convergence of $\{(x_k, \lambda_k)\}$ and $\{x_k\}$ but not $\{\lambda_k\}$; while a weakly x-dominated update leads to the q-linear convergence of $\{(x_k, \lambda_{k+1})\}$ but not $\{x_k\}$ or $\{\lambda_k\}$. Of course in both cases we will have r-linear convergence of $\{x_k\}$ and $\{\lambda_k\}$.

4. A Demonstration of the Convergence Theory. In this section we apply the convergence theory developed in Section 3 to the four standard multiplier updates given by (1.9)-(1.12). As stated in the introduction we do this more as a test and demonstration of the unified theory than as a statement about the updates themselves.

In what follows we will be requiring B to be of bounded deterioration at A^\sharp with respect to the particular multiplier update in question. Implicit in the work of Han [26], Tapia [39], Glad [25] and Boggs, Tolle and Wang [5] is a proof that the Broyden and PSB secant updates and the DFP and the BFGS secant updates in the case of positive definite A^\sharp are of bounded deterioration for any particular update. See in particular the comments in the proof of our Corollary 5.5.

PROPOSITION 4.1. *Given $r \in (0,1)$ there exists $c(r) \geq 0$ such that for each $c > c(r)$ we can find a neighborhood of $(x_*, \lambda_*, A^\sharp)$ so that in this neighborhood the projection multiplier update (1.10) is x-dominated with constant $\phi(c) < r$. Hence the DMM using this multiplier update and $c \geq c(r)$ is locally convergent and satisfies (3.1).*

Proof. We will give the proof for any consistent update which does not depend on c or λ and has continuous partial derivative with respect to x , since it is essentially the

same as the proof for the special case (1.10).

Choose $c(r) > \bar{c}$ in Lemma 2.1 so that A^\sharp is invertible. By consistency and the mean-value theorem we have

$$\begin{aligned} |(A^\sharp)^{-1} \nabla g_\bullet (\lambda_+ - \lambda_\bullet)| &= |(A^\sharp)^{-1} \nabla g_\bullet [U(x, B) - U(x_\bullet, B)]| \\ &\leq |(A^\sharp)^{-1} \nabla g_\bullet | \nabla_x U(x_\bullet + \theta(x - x_\bullet), B) | |x - x_\bullet| \end{aligned} \quad (4.1)$$

for some $\theta \in (0, 1)$. The proof now follows from (2.2.e) of Lemma 2.1 and the continuity of $\nabla_x U(x, B)$. \square

A proposition exactly like Proposition 4.1 can be proved for the Newton multiplier update (1.12). However, the Newton update does not require large c . In fact we have arbitrary good local linear convergence with $c = 0$. Recall that A_\bullet denotes $\nabla_x^2 l(x_\bullet, \lambda_\bullet)$.

PROPOSITION 4.2. *Given $r \in (0, 1)$ there exists a neighborhood of $(x_\bullet, \lambda_\bullet, A_\bullet)$ such that the Newton multiplier update (1.12) is x -dominated with constant $\phi < r$. Hence the DMM using this multiplier update and $c = 0$ is locally convergent and satisfies (3.1).*

Proof. A straightforward calculation shows that $U(x, \lambda, B)$ as given by (1.12) is consistent and independent of λ . So we can write

$$\lambda_+ - \lambda_\bullet = U(x, \lambda, B) - \lambda_\bullet = U(x, \lambda_\bullet, B) - U(x_\bullet, \lambda_\bullet, B). \quad (4.2)$$

Differentiating $U(x, \lambda_\bullet, B)$ with respect to x at $x = x_\bullet$ gives

$$\nabla_x U(x_\bullet, \lambda_\bullet, B) = (\nabla g_\bullet^\sharp B^{-1} \nabla g_\bullet)^{-1} \nabla g_\bullet^\sharp [I - B^{-1} A_\bullet]. \quad (4.3)$$

From (4.3) we see that $\nabla_x U(x_\bullet, \lambda_\bullet, A_\bullet) = 0$. Using the mean-value theorem, (4.2) and (4.3) we obtain

$$|\lambda_+ - \lambda_\bullet| \leq |\nabla_x U(x_\bullet + \theta(x - x_\bullet), \lambda_\bullet, B)| |x - x_\bullet| \quad (4.4)$$

for some $\theta \in (0, 1)$. The proposition now follows by observing that by continuity the derivative term in (4.4) can be made arbitrarily small for x near x_\bullet and B near A_\bullet . \square

PROPOSITION 4.3. *Given $r \in (0, 1)$ there exists $c(r) \geq 0$ such that for each $c > c(r)$ we can find a neighborhood of $(x_\bullet, \lambda_\bullet, A^\sharp)$ so that in this neighborhood the Hestenes-Powell multiplier update (1.9) is weakly x -dominated with constant $\phi(c) < r$. Hence the DMM using this multiplier update and $c \geq c(r)$ is locally convergent and satisfies (3.15).*

Proof. Consider U given by (1.9). Write $\lambda_{++} = U(x_+, \lambda_+, B_+)$. For $i = 1, \dots, m$ there exist $\theta_i \in (0, 1)$ such that

$$g_i(x_+) - g_i(x_\bullet) = \nabla g_i(x_\bullet + \theta_i(x_+ - x_\bullet))^t (x_+ - x_\bullet).$$

Let

$$\nabla g_\theta = [\nabla g_1(x_\bullet + \theta_1(x_+ - x_\bullet)), \dots, \nabla g_m(x_\bullet + \theta_m(x_+ - x_\bullet))],$$

so that $g(x_+) - g(x_\bullet) = \nabla g_\theta^t (x_+ - x_\bullet)$. We can now write, since $g(x_\bullet) = 0$,

$$\lambda_{++} - \lambda_* = \lambda_+ - \lambda_* + c \left[g(x_+) - g(x_*) \right] = \lambda_+ - \lambda_* + c \nabla g^t(x_+) (x_+ - x_*). \quad (4.5)$$

A straightforward calculation shows that

$$\begin{aligned} x_+ - x_* &= \left[I - B^{-1} A^t \right] (x - x_*) - B^{-1} \nabla g(x) (\lambda_+ - \lambda_*) \\ &\quad - B^{-1} \left[\nabla_x L(x, \lambda_*, c) - \nabla_x L(x_*, \lambda_*, c) - A^t (x - x_*) \right] \end{aligned} \quad (4.6)$$

Now using Lemma 2.2 with $F(x) = \nabla_x L(x, \lambda_*, c)$, $u = x_*$ and $v = x$, and combining (4.5) with (4.6) we obtain K_L such that

$$\begin{aligned} |\lambda_{++} - \lambda_*| &\leq |I - c \nabla g^t B^{-1} \nabla g(x)| |\lambda_+ - \lambda_*| \\ &\quad + c |\nabla g^t| \left[|I - B^{-1} A^t| + K_L |B^{-1}| |x - x_*| \right] |x - x_*|. \end{aligned} \quad (4.7)$$

The proposition now follows from (4.7) by first choosing $c(r)$ large guided by (2.2.d) of Lemma 2.1. Then choosing (x, B) close to (x_*, A^t) and if needed further restricting the choice of x . \square

PROPOSITION 4.4. *Given $r \in (0, 1)$ there exists $c(r) \geq 0$ such that for each $c > c(r)$ we can find a neighborhood of (x_*, λ_*, A^t) so that in this neighborhood the Buys multiplier update (1.11) is weakly x -dominated with constant $\phi(c) < r$. Hence the DMM using this multiplier update and $c \geq c(r)$ is locally convergent and satisfies (3.15).*

Proof. Exactly the same argument used in the proof of Proposition 4.3 will lead us to an expression of the form (4.7) with the factor c replaced by $(\nabla g(x_+)^t B^{-1} \nabla g(x_+))^{-1}$. The proof now follows by choosing $c(r)$ large enough so that $|(A^t)^{-1} \nabla g_*| < r$ and then choosing (x, B) sufficiently close to (x_*, A^t) so that the appropriate factors will lead to a $\phi(c)$ which is less than 1. \square

The proofs given above suggest that for the Hestenes-Powell update c must be very large, for the projection update and the Buys update c should be of the same order and need not be particularly large and for the Newton update $c = 0$ works fine.

5. Superlinear convergence. In this section we will develop a theory for studying the q-superlinear convergence of the sequence $\{x_k\}$ generated by the DMM. This theory is closely related to the theory for unconstrained optimization. Recall that the sequence $\{x_k\}$ is q-superlinearly convergent to x_* if

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = 0. \quad (5.1)$$

According to the Dennis-Moré [17] characterization theory for quasi-Newton methods for unconstrained optimization in the case of a quasi-Newton method applied to the unconstrained minimization of the functional $L(x, \lambda_*, c)$ (equivalently the idealized DMM where the choice for λ_k is λ_*), a necessary and sufficient condition for q-superlinear

convergence is

$$\lim_{k \rightarrow \infty} \frac{|(B_k - A^c)s_k|}{|s_k|} = 0, \quad (5.2)$$

assuming convergence of the iterates.

It seems reasonable that (5.2) will play an important role in our characterization of q-superlinear convergence in constrained optimization and that by itself will not imply q-superlinear convergence. In fact, it is not surprising that the additional condition that is needed is that the multipliers λ_k converge to λ_* sufficiently fast, i.e.

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|s_k|} = 0. \quad (5.3)$$

Recall that the sequences $\{x_k\}$ and $\{\lambda_k\}$ are generated by the DMM (1.3)-(1.6). In addition to assumptions A1-A4 we will assume

A5. The iterates $x_k \in D$, and $\lim_{k \rightarrow \infty} x_k = x_*$.

The following is our first characterization of q-superlinear convergence of the sequence $\{x_k\}$.

THEOREM 5.1. *Any two of (5.1), (5.2), or (5.3) imply the third.*

Proof. From (1.4) we can write

$$\begin{aligned} -\nabla_x L(x_{k+1}, \lambda_*, c) &= -[\nabla_x L(x_{k+1}, \lambda_*, c) - \nabla_x L(x_k, \lambda_*, c) - A^c s_k] \\ &\quad + \nabla g_k(\lambda_{k+1} - \lambda_*) + (B_k - A^c) s_k. \end{aligned} \quad (5.4)$$

Let $F(x) = \nabla_x L(x, \lambda_*, c)$. Then $F(x_*) = 0$, and $F'(x) = A^c$. From Lemma 2.2 there exist positive β and α such that

$$\beta |x_{k+1} - x_*| \geq |\nabla_x L(x_{k+1}, \lambda_*, c)| \geq \alpha |x_{k+1} - x_*| \quad (5.5)$$

for k sufficiently large. Also from Lemma 2.3 there exist a positive constant K_1 such that

$$|\nabla_x L(x_{k+1}, \lambda_*, c) - \nabla_x L(x_k, \lambda_*, c) - A^c s_k| \leq K_1 \sigma(x_k, x_{k+1}) |s_k|. \quad (5.6)$$

Now by dividing (5.4) by $|s_k|$ and observing (5.6) we see that any two of the following three statements implies the third

$$\lim_{k \rightarrow \infty} \frac{|\nabla_x L(x_{k+1}, \lambda_*, c)|}{|s_k|} = 0 \quad (5.7)$$

$$\lim_{k \rightarrow \infty} \frac{|\nabla g_k(\lambda_{k+1} - \lambda_*)|}{|s_k|} = 0 \quad (5.8)$$

$$\lim_{k \rightarrow \infty} \frac{|(B_k - A^*)\delta_k|}{|\delta_k|} = 0. \quad (5.9)$$

Inequality (5.5) shows that (5.7) is equivalent to

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x_*|}{|\delta_k|} = 0. \quad (5.10)$$

A fairly straightforward argument using the triangle inequality can be used to show that (5.1) is equivalent to (5.10). Lemma 2.4 shows that (5.8) is equivalent to (5.3). Finally, (5.9) is exactly (5.2). \square

Our second characterization of q-superlinear convergence of the sequence $\{x_k\}$ will use the projection operator onto the tangent space of the constraints, i.e.

$$P(x) = I - \nabla g(x) [\nabla g^t(x) \nabla g(x)]^{-1} \nabla g^t(x).$$

Let $P_* = P(x_*)$, and $P_k = P(x_k)$. Before stating the next theorem we need a technical result.

LEMMA 5.2. Let $H_c : R^n \rightarrow R^n$ be a function defined by

$$H_c(x) = P(x) \nabla_x l(x, \lambda_*) + c \nabla g_* g(x)$$

for $x \in D$. Then H_c is continuously differentiable in D ,

$$H_c(x_*) = 0, \quad (5.11)$$

and

$$H_c'(x_*) = P_* A_* + c \nabla g_* \nabla g_*^t. \quad (5.12)$$

Moreover, if $c \neq 0$ then $H_c'(x_*)$ is nonsingular.

Proof. From assumptions A1-A4 we have that H_c is continuously differentiable in D , and (5.11) holds. By differentiating H_c and evaluating H_c' at x_* we get (5.12). The nonsingularity of $H_c'(x_*)$ is due to the following fact. Let $d \neq 0$ and

$$(P_* A_* + c \nabla g_* \nabla g_*^t) d = 0.$$

Since ∇g_* has full column rank we have

$$P_* A_* d = 0 \quad \text{and} \quad \nabla g_*^t d = 0.$$

Therefore, $P_* d = d$ and $d^t A_* d = 0$, which is a contradiction since $d^t A_* d > 0$ for all d such that $\nabla g_*^t d = 0$. \square

THEOREM 5.3. A necessary and sufficient condition for $\{x_k\}$ to converge q-superlinearly to x_* is

$$\lim_{k \rightarrow \infty} \frac{|P_k (B_k - A_*) s_k|}{|s_k|} = 0, \quad (5.13)$$

and

$$\lim_{k \rightarrow \infty} \frac{|g_k + \nabla g_k^t s_k|}{|s_k|} = 0. \quad (5.14)$$

Proof. From (1.4) we have

$$B_k s_k + \nabla_x L(x_k, \lambda_{k+1}, c) = 0$$

hence

$$-\nabla_x L(x_k, \lambda_{k+1}, c) = B_k s_k + \nabla g_k(\lambda_{k+1} - \lambda_*).$$

Multiplying our last expression by P_k and observing that $P_k \nabla g_k = 0$ we obtain

$$P_k B_k s_k + P_k \nabla_x l(x_k, \lambda_*) = 0.$$

Adding on both sides $-H_c(x_{k+1})$ yields

$$-H_c(x_{k+1}) = P_k B_k s_k - [H_c(x_{k+1}) - H_c(x_k)] - c \nabla g_* g_k.$$

Using (5.12) we obtain

$$\begin{aligned} -H_c(x_{k+1}) &= -[H_c(x_{k+1}) - H_c(x_k) - H_c'(x_*) s_k] \\ &\quad - c \nabla g_* [g_k + \nabla g_k^t s_k] \\ &\quad + (P_k - P_*) A_* s_k \\ &\quad + P_k (B_k - A_*) s_k. \end{aligned} \quad (5.15)$$

From Lemma 2.1, and Lemma 5.2 there exist for $c \neq 0$, $\beta \geq \alpha > 0$ such that

$$\beta |x_{k+1} - x_*| \geq |H_c(x_{k+1})| \geq \alpha |x_{k+1} - x_*| \quad (5.16)$$

for k sufficiently large. Also from Lemma 2.3 there exists a positive K_1 such that

$$|H_c(x_{k+1}) - H_c(x_k) - H_c'(x_*) s_k| \leq K_1 \sigma(x_k, x_{k+1}) |s_k|. \quad (5.17)$$

Consider the condition

$$\lim_{k \rightarrow \infty} \frac{|H_c(x_{k+1})|}{|s_k|} = 0. \quad (5.18)$$

By (5.16) we see that (5.18) is equivalent to

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x_*|}{|s_k|} = 0 \quad (5.19)$$

which (as argued in the proof of Theorem 5.1) in turn is equivalent to (5.1) and q-superlinear convergence.

Assume (5.13) and (5.14). Divide (5.15) by $|s_k|$, recall (5.17) and take limits to obtain (5.18). This establishes the q-superlinear convergence of $\{x_k\}$ to x_* .

Let us now assume (5.1). We first argue that this assumption implies (5.14). To see this we write

$$g_k + \nabla g_k^t s_k = -(g_{k+1} - g_k - \nabla g_k^t s_k) + (g_{k+1} - g_k). \quad (5.20)$$

Dividing (5.20) by $|s_k|$, calling on Lemma 2.2 to bound the first term on the right-hand side of (5.20) and the mean-value theorem to bound the second term we have

$$\lim_{k \rightarrow \infty} \frac{|g_k + \nabla g_k^t s_k|}{|s_k|} \leq K \lim_{k \rightarrow \infty} \frac{|x_{k+1} - x_*|}{|s_k|} \quad (5.21)$$

for some positive constant K . If now recall the fact that (5.1) is equivalent to (5.19) we see that (5.21) implies (5.14).

We are assuming (5.1) holds or equivalently (5.18). We have established (5.14). This means that if we divide (5.15) by $|s_k|$ and recall (5.14), (5.17) and (5.18) when taking limits we obtain (5.13). \square

It is of considerable interest to see what Theorem 5.3 gives when the multiplier update is the Newton update (1.12). Recall that in this case the DMM is equivalent to Successive Quadratic Programming. Theorem 10.2 of Tapia [40] says that the DMM using the Newton multiplier update satisfies linearized constraints, i.e., (5.14) holds. This gives us the following corollary to Theorem 5.3.

COROLLARY 5.4. *Let the sequences $\{x_k\}$ be generated using the DMM (1.9)-(1.6) with the Newton multiplier update formula (1.12). Then a necessary and sufficient condition for $\{x_k\}$ to converge q-superlinearly to x_* is*

$$\lim_{k \rightarrow \infty} \frac{|P_k(B_k - A_*)s_k|}{|s_k|} = 0 \quad (5.22)$$

Corollary 5.4 is the Boggs, Tolle and Wang [5] characterization theorem discussed in Section 1.

The following Corollary says that the most popular secant updates, Broyden, PSB, DFP and BFGS, which are known to give local q-superlinear convergence in the case of unconstrained optimization give local q-superlinear convergence in x in the case of constrained optimization provided one uses the Newton multiplier update.

By using an obvious weighting and the infinity norm it is not difficult to see that if one obtains q-superlinear convergence in x using an x -dominated multiplier update, then q-superlinear convergence in the pair (x, λ) follows. The converse is not necessarily true.

The fact that the DMM using the Newton update, equivalently SQP, gives q-superlinear convergence in the pair (x, λ) was established by Han [26], Tapia [39] and Glad [25]. That one also obtains q-superlinear convergence in x alone was established by Boggs, Tolle and Wang [5].

COROLLARY 5.5. *Consider the DMM using the updates U and B where U is x -dominated and B is either the PSB, the DFP or the BFGS secant update. Assume that in*

the case of DFP and BFGS secant updates the matrix A^ε is positive definite. Then the DMM is locally q -linearly convergent in x . Moreover, if U is the Newton multiplier update then we also have q -superlinear convergence in x .

Proof. Implicit in the works of Han [26], Glad [25], Tapia [39] and Boggs, Tolle and Wang [5] for the Broyden, PSB and DFP secant updates is an inequality of the form

$$\begin{aligned} \|B_{k+1} - A^\varepsilon\| &\leq \left[1 + \alpha_1 \sigma(x_k, x_{k+1})\right] \|B_k - A^\varepsilon\| \\ &\quad + \beta_1 \sigma(x_k, x_{k+1}) + \beta_2 |\lambda_{k+1} - \lambda_*| \end{aligned} \quad (5.23)$$

where $\alpha_1 > 0$, $\beta_1 > 0$ and $\beta_2 > 0$ and as before

$$\sigma(x_k, x_{k+1}) = \max\{|x_{k+1} - x_*|, |x_k - x_*|\}.$$

Now using the assumption that the multiplier update U is x -dominated we can write (5.23) as

$$\begin{aligned} \|B_{k+1} - A^\varepsilon\| &\leq \left[1 + \alpha_1 \sigma(x_k, x_{k+1})\right] \|B_k - A^\varepsilon\| \\ &\quad + \alpha_1 \sigma(x_k, x_{k+1}). \end{aligned} \quad (5.24)$$

From Proposition 4.2 we have local q -linear convergence. Now, an argument identical to the one used by Broyden, Dennis and Moré [6] can be used to establish

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - A^\varepsilon)s_k\|}{\|s_k\|} = 0. \quad (5.25)$$

By observing that $P_k \nabla g_k = 0$ and $\|P_k\| = 1$ we get

$$\|P_k (B_k - A_*) s_k\| \leq \|P_k (B_k - A^\varepsilon) s_k\| + c \|\nabla g_* \nabla g_k^\varepsilon - \nabla g_k \nabla g_k^\varepsilon\| \|s_k\|. \quad (5.26)$$

Dividing (5.26) by $\|s_k\|$, taking limits and using (5.25) we get (5.22).

The proof for the BFGS update requires one to work with the inverse update. The proof is then essentially the same as that for the DFP (see Broyden, Dennis and Moré [6]). \square

It is of interest to emphasize that the DMM using the projection multiplier update and the secant updates listed above satisfies (5.25) but does not lead to superlinear convergence. This fact should enhance the appreciation for Theorem 5.3.

Acknowledgment. A portion of this work is contained in the first author's doctoral thesis under the supervision of the second two authors in the Department of Mathematical Sciences, Rice University. The first author would like to thank the Venezuelan government for support during his graduate studies.

Quasi-Newton Methods For Constrained Optimization

The authors gratefully thank John Dennis Jr. for numerous helpful suggestions and discussions. They also sincerely thank Paul Boggs and Michael Overton for reading an earlier draft of the paper and making numerous suggestions and comments which led to a substantial revision and improvement of the original paper.

REFERENCES

- [1] M. Bertochi, E. Cavalli and E. Spedicato, *Computational performance of diagonalized multiplier quasi-Newton methods for nonlinear optimization with equality constraints*, in Numerical Optimization of Dynamic Systems, 1, C.W. Dixon and G.P. Szego eds., North Holland, Amsterdam, (1979), pp. 247-268.
- [2] Y. Bard and J. Greenstadt, *A modified Newton method for optimization with equality constraints*, Optimization, R. Fletcher, ed., Academic press, New York, (1969).
- [3] D.P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic Press, New York, (1982).
- [4] M.C. Biggs, *On the convergence of some constrained minimization algorithms based on Recursive Quadratic Programming*, J.Inst.Math.Appl., 21, (1978), pp. 67-81.
- [5] P.T. Boggs, J.W. Tolle and P. Wang, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control and Optimization, 20, (1982), pp. 161-171.
- [6] C.G. Broyden, J.E. Dennis Jr. and J.J. Moré, *On the local and superlinear convergence of quasi-Newton methods*, J.Inst.Math.Appl., 12, (1973), pp. 223-245.
- [7] J.D. Buys, *Dual algorithms for constrained optimization problems*, Ph.D. Thesis, University of Leiden, Netherlands.
- [8] J.D. Buys and P.C. Haarhoff, *A new method for the optimization of a nonlinear function subject to nonlinear constraints*, Computer Journal, 13, (1970), pp. 178-184.
- [9] R.H. Byrd, *Local convergence of the diagonalized method of multipliers*, J.O.T.A., 26, (1978), pp. 483-497.
- [10] R.H. Byrd, *On the convergence of constrained optimization methods with accurate Hessian information on a subspace*, TR. CU-CS-270-84, Dept. Computer Science, Univ. of Colorado, Boulder, CO., (1984).
- [11] M.R. Celis, J.E. Dennis Jr. and R.A. Tapia, *A trust region strategy for nonlinear equality constrained optimization*, Numerical Optimization 1984, P.T. Boggs, R.H. Byrd and R.B. Schnabel, eds., SIAM, Philadelphia, (1985), pp. 71-82.
- [12] T.F. Coleman and A.R. Conn, *Nonlinear programming via an exact penalty function: Asymptotic analysis*, Math. Prog., 24, (1982), pp. 123-136.

- [13] T.F. Coleman and A.R. Conn, *Nonlinear programming via an exact penalty function: Global analysis*, Math. Prog., 24, (1982), pp. 137-161.
- [14] T.F. Coleman and A.R. Conn, *On the local convergence of a quasi-Newton method for the nonlinear programming problem*, SIAM J. Numer. Anal., 21, (1984), pp. 755-769.
- [15] R.S. Dembo, S.C. Eisenstat and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19, (1982), pp. 400-408.
- [16] J.E. Dennis, *Toward a unified convergence theory for Newton-like methods*, in Non-linear Functional Analysis and Applications, L.B. Rall eds., Academic Press, New York, (1971), pp. 425-472.
- [17] J.E. Dennis Jr. and J.J. Moré, *A characterization of q -superlinear convergence and its applications to quasi-Newton methods*, Math. Comp., 28, (1974), pp. 549-560.
- [18] J.E. Dennis Jr. and R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall, Englewood Cliffs, NJ, (1983).
- [19] J.E. Dennis Jr. and H.F. Walker, *Local convergence theorems for quasi-Newton methods*, SIAM J. Numer. Anal., 18, (1981), pp. 949-987.
- [20] R. Fletcher, *Practical Methods of Optimization*, Volume 2, Constrained Optimization, John Wiley, New York, (1981).
- [21] R. Fontecilla, *A general convergence theory for quasi-Newton methods for constrained optimization*, Ph.D. Thesis, Math. Sci. Dept., Rice University, (1983).
- [22] R. Fontecilla, *The lack of positive definiteness in the Hessian in constrained optimization*, TR. 1334, Dept. Computer Science, Univ. of Maryland, College Park, Md., (1983).
- [23] U.M. Garcia-Palomares and O.L. Mangasarian, *Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems*, Math. Prog., 11, (1976), 1-13.
- [24] P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, New York, (1981).
- [25] S.T. Glad, *Properties of updating methods for the multipliers in augmented Lagrangians*, J.O.T.A., 28, (1979), pp. 135-156.
- [26] S.P. Han, *Superlinearly convergent variable metric algorithms for general nonlinear programming problems*, Math. Prog., 11, (1976), pp. 263-282.

- [27] S.P. Han, *Dual variable metric algorithms for constrained optimization*, SIAM J. Control and Optimization, 15, (1977), pp. 546-565.
- [28] M.R. Hestenes, *Multiplier and gradient methods*, J.O.T.A., 4, (1969), pp. 303-320.
- [29] A. Miele, E.E. Cragg, R.B. Iyer and A.V. Levy, *Use of the augmented penalty function in mathematical programming problems*, Part I, J.O.T.A., 8, (1971), pp. 151-180.
- [30] A. Miele, P.E. Moseley, A.V. Levy and G.M. Coggins, *On the method of multipliers for mathematical programming problems*, J.O.T.A., 10, (1972), pp. 1-33.
- [31] J. Nocedal and M. Overton, *Projected Hessian updating algorithms for nonlinearly constrained optimization*, TR. 95, Dept. Computer Science, New York Univ., New York, N.Y., (1983).
- [32] J.M Ortega and W.C. Rheinboldt, *Iterative solutions of nonlinear equations in several variables*, Academic Press, New York, (1970).
- [33] M.J.D. Powell, *A method for nonlinear constraints in minimization problems*, in Optimization, R. Fletcher ed., Academic Press, New York, (1969), pp. 283-298.
- [34] M.J.D. Powell, *The convergence of variable metric methods for nonlinearly constrained optimization calculations*, Nonlinear Programming 3, Mangasarian, Meyer and Robinson, eds., Academic Press, New York, (1978), pp. 27-63.
- [35] M.J.D. Powell, *Variable metric methods for constrained optimization*, Report DAMTP 82/NA 5, Univ. of Cambridge, England, (1982).
- [36] A. Stachurski, *Superlinear convergence of Broyden's bounded θ -class of methods*, Math. Prog., 20, (1981), pp. 196-212.
- [37] R. A. Tapia, *Newton's method for optimization problems with equality constraints*, SIAM J. Numer. Anal., 11, (1974), pp. 874-886.
- [38] R. A. Tapia, *A stable approach to Newton's method for general mathematical programming problems in R^n* , J.O.T.A., 14, (1974), pp. 453-476.
- [39] R. A. Tapia, *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization*, J.O.T.A., 22, (1977), pp. 135-194.
- [40] R.A. Tapia, *Quasi-Newton methods for equality constrained optimization: equivalence of existing methods and a new implementation*, Nonlinear Programming 3, Mangasarian, Meyer and Robinson, eds., Academic Press, New York, 1978, pp. 125-164.

- [41] R.B. Wilson, *A simplicial algorithm for concave programming*, Ph. D. Thesis, Harvard University (1963).